The emergence of information in biology: common principles in the evolution of organization

Daniel Thomas Wright
Graduate School of Library and Information Science
University of Illinois at Urbana-Champaign
501 E. Daniel Street, Champaign, IL 61820

ABSTRACT

Information organization and management are commonly understood to be the products of conscious processes undertaken by human beings. While not incorrect, this view is narrow and excludes a wide range of natural, evolved informational mechanisms in biological systems. The creation of human information organization systems that can exercise a degree of autonomy in interactions with the world – so that they can be flexible and adapt to changing conditions – would benefit from an understanding of the biological information systems that do this. By broadening our view of information organization to include these natural systems, we stand to gain a deeper understanding of both principles of information organization and the natural world.

1. INTRODUCTION

In LIS, systems of information organization are human designed and human implemented, and are often understood to be synonymous with classification systems. This is true across all sorts of technologies, from classical library classifications like the Decimal Classification (DC), to new systems like user-driven tagging. All these systems of information organization have a historical and collective developmental trajectory. Likewise, the informational systems of organisms, such as the genetic code and neurological systems, are evolved systems and as such have a historical, developmental trajectory. Understanding the evolutionary principles and possible trajectories of these systems will contribute to our understanding of human designed systems by presenting a much more general picture of the essential features of functional information systems.

Human designed, artificial information organization systems are a product of the underlying human biological constraints, and depend on language and culture for their existence. They have varying levels of rigidity, from the formal specifications of DC systems to the more organic, evolutionary nature of folksonomies and on-line user tagging-based collections. While the latter seem to have more obviously in common with the biological systems described above, in that they demonstrate some of the same self-organizing evolutionary properties, they are not founded on an understanding of the underlying principles and have been criticized as lacking real-world utility. The contrasts between these and the formal systems may be informative; and in the same way, the contrasts and similarities between all of these may reveal fundamental constraints and principles around the whole space of solutions to the information organization problem.

We have observed the following properties of information organization systems, and take them as assumptions grounding the discussion that follows. First, it is clear that the organization of information is inherently a balance between dynamic and stable properties, in that an organization system needs to be able to accommodate innovation and reflect change in its environment, while still having enduring utility. Second, information organization appears in many different contexts: at a biological level, organisms use organization to real, adaptive, functional advantage; formal classifications are used throughout LIS; and informal information organization systems are found everywhere around us. Third, information organization is clearly an *emergent* phenomena – it hasn't "always been there", but has emerged many times in many contexts, ranging from evolved organization in biology to engineered systems in LIS. In all cases, organization emerges in response to some need, and emerges as a collective phenomena. Finally, evidence from evolutionary genetics and from investigations of abstract models of language strongly suggests that information organizations of high dynamic complexity and robustness can evolve naturally – without planning or intent.

2. WHAT IS BIOLOGICAL INFORMATION ORGANIZATION?

Examples of biological information organization include the most fundamental cellular systems like the organization of the genome, to the most complex products of biological information – those of language and culture [1]. Biological information self-organizes in many ways. These systems of organized information have been implicated as the crucial developments in the evolution of higher organisms capable of coordinated action and complex communications [4].

In the genome, organization is exhibited in gene ordering, clustering of related genes, and the structure of coding and

[1] By which I simply mean that culture and language are necessarily constrained by, and outcomes of, human biology.

non-coding material in the genome. These properties exist for a variety of functional reasons; for example, co-located genes are able to be processed and expressed more rapidly by the cell. Genome organizations have great consistencies across a wide range of organismal classes [1], indicating that they are functionally important, and were a relatively early evolutionary development.

Two of the most significant products of biological complexity are human language and culture. Language may be thought of as a structured system of information, used both for communication and the organization of information. It is the product of certain aspects of biological evolution, like the human brain and complex vocal apparatus. Additionally, language itself is a constantly evolving communal phenomena, and the communicative and organizational aspects of language are mutually reinforcing and constantly evolving. It exhibits its own characteristic evolutionary dynamics, while having much in common with other aspects of biological organization. Language is necessary for the existence of culture and society, and enables the development and specification our artificial systems of organization and classification.

3. INFORMATION EMERGENCE

The evolution of informational systems in biology is a problem of emergent information and semantics: starting from abiotic chemistry, how did we get to having representational and communicative information? At the molecular level, one of the earliest and most obvious examples is the question of the evolution of the genetic code. The range of emergent information systems spans all areas of biology, from the earliest molecular systems through modern human language. A question of great interest in this context is: what are the common principles behind the evolution of these systems of information representation and organization?

One way of getting at these common principles is to build abstractions from their essential features for computer modeling. To this end, we can develop simplified model organisms, and investigate ways that they are able to bootstrap their own semantics and information structures. We know that both organisms and human-developed organization systems must be able to adapt to changing demands throughout their life-cycles; the problem of engineering the accommodation of all contingencies into systems at their inception is clearly intractable [2], so instead we want to address this problem by creating systems able to learn and adapt. Investigations into the evolution of artificial languages have demonstrated that relatively complex representational and semantic systems can be evolved autonomously [3]. This suggests that it is reasonable to think that building software systems able to evolve and use their own organizational principles and content is a practical goal.

4. CONCLUSION

The creation of human information organization systems that can exercise a degree of autonomy in their interaction with the world – so that they can be flexible and adapt to changing conditions – is an enterprise that will benefit from an understanding of principles of the biological information systems that are *already* able to do this. Research into these principles is required, as we currently have only a very high-level, abstract picture of what they might be. We plan to attack this problem, beginning with a theoretical account and computer model of the emergence of heritary information in primitive biology. Elaborating the essential, abstract features of biological self-organizing information systems is a required first step in modeling. Our goal is to do this, to contribute to our understanding of the fundamental principles behind information organization.

5. REFERENCES

[1] E. V. Koonin and Y. I. Wolf. Genomics of bacteria and archaea: the emerging dynamic view of the prokaryotic world. *Nucleic Acids Research*, 36(21):6688–6719, 2008.

[2] H. A. Olson. Mapping beyond dewey's boundaries: Constructing classificatory space for marginalized knowledge domains. *Library Trends*, 47(2):233–254, 1998.

[3] S. Swarup and L. Gasser. Simple, but not too simple: Learnability vs. functionality in language evolution. In *The Seventh Evolution of Language Conference (EvoLang'08)*, 11–15 March 2008.

[4] E. Szathmary and J. Maynard-Smith. The major evolutionary transitions. *Nature*, 374:227–232, 16 March 1995.